

## FFCUL/LASIGE TEAM THIRD YEAR INTERIM REPORT

The Epiwork project started in February 2009 and will run for 48 months. According to the work plan FFCUL is involved in the following work packages:

- WP 1 Population Models and Contact Networks
- WP 3 Information platform
- WP 4 Epidemic Modelling Platform
- WP 7 Management

The FFCUL participation in Epiwork involves two groups:

- CMAF Group
- LASIGE Group.

This report summarises the progress in the fifth semester of the project activity by the LASIGE Group, which leads WP3 and participates in WP4 and WP7.

In the reporting period (1<sup>st</sup> semester of the 3<sup>rd</sup> year), the commitment was 27.51 persons.month, 21.31 pm from technicians, and 6.20 pm from permanent staff.

The LASIGE Team working on WP3 and WP4 in this period includes:

- Mário J. Silva (FCUL and IST Faculty, worked in the reporting period in WP3, WP4 and WP7)
- Francisco Couto (FCUL Faculty, worked in the reporting period in WP3)
- Dulce Domingos (FCUL Faculty, worked in the reporting period in WP3 and WP4)
- Carla Patrícia Sousa (Graduate student, worked full-time in the reporting period, left on July 7th).
- João Zamite (Graduate student, worked full-time in the reporting period).
- Juliana Duque (Graduate student, worked full-time in the reporting period).
- João Ferreira (Graduate student, worked half-time until October and full-time since then).
- Fernando Silva (Software developer hired for 4 months between March and July)
- Paulo Graça (software developer, is working full-time since December)
- Christian Goellner (undergraduate student, worked full-time in July and quit at the end of that month)
- Carlos Sousa (undergraduate student, is working full-time since July)
- Sheliza Fajal (undergraduate student, worked full-time since July, but quit in October)
- Tiago Posse Sousa (undergraduate student, is working full-time since August)

The following changes occurred in the FFCUL/LASIGE Team:

**Mário J. Silva** left the University Lisbon to join Instituto Superior Técnico, IST, also in Lisbon. However, he remains, as an invited researcher at LASIGE, the leader of the FFCUL participation in EPIWORK.

There is a discrepancy in the profiles of the people hired to work on Epiwork at LASIGE and the initially planned, the reason being that we were unable to fill the software engineer and a post-doc positions as planned, with previous directly related experience and salaries matching our budget.

As a result, we started using Informatics Engineering and Biomedical Informatics students on the infrastructure setup assisted by FCUL faculty and senior technical staff. LASIGE has hired more people in the first two years at a reduced cost, which was the contingency measure found to manage the situation.

**Carla Patrícia Sousa** and **Juliana Duque** completed their Master degrees and quit the project in July and December, respectively.

**Fernando Silva** is an expert in Drupal development who worked part-time for the project from April to July on the customization of this content management system for the Epidemic Marketplace.

**Paulo Graça** was recruited in December 2011 for working full-time on the project as senior Drupal programmer and software development coordination.

**Christian Goellner** and **Sheliza Fajal** were new junior researchers who joined the project on the 1<sup>st</sup> of July to work in the development of the Epidemic Marketplace user interface. Christian Goellner quit at the end of July and Sheliza Fajal quit in October.

**Carlos Sousa** is a new junior researcher who joined the project on the 1<sup>st</sup> of July and has been working in the development of the Epidemic Marketplace user interface.

**Tiago Posse** is a new junior researcher who joined the project on the 1<sup>st</sup> of August and is working in the support of the base infrastructure and helping the development of the Epidemic Marketplace user interface.

In addition, in the period we also had the collaboration of **Cátia Pesquita**, a graduate research student in the research team, on the production of the Deliverable D3.5. Cátia has expertise on ontologies and participated in the discussions for selecting the ontologies included in NERO and reviewed the text of the deliverable.

## 1. WP3 — Information platform: Work progress, collaborations and achievements during the period

This Work Package is lead by FFCUL, with a total contribution of 82 persons.month (60 hired technicians + (22) academics).

### 2.1 Background Information on WP3

Work package number	3	Start date or starting event:					Month 1	
Work package title	Information platform							
Activity type	RTD							
Participant number	12	1	2	3	4	5	9	10

Participant short name	FFCUL	ISI	FGC-IGC	TAU	MPI-DS	AIBV	BIU	FBK-IRST
Person-months per participant	82	60	4	8	42	2	11	6

The whole WP3 activity is structured into four tasks:

### **Task 3.1 – Data Collection.**

**Participants:** FFCUL, ISI, FBK-IRST, BIU, MPI-DS, FGC-IGC, AIBV.

**Description:** Realistic simulations of epidemic processes crucially depend on the availability of datasets describing human behaviour and pathogen-host interactions. Datasets include population movement data, social and behavioural data, health related data, geographic data, detailed geo-temporal epidemic incidence and immunization data, pathogen evolution and multi-strains circulation data. Data can come from a variety of different sources, including hospital records, country statistics, Web content, and others. It can range from a global scale, such as the worldwide air transportation infrastructure, down to the detailed description of individual **activities** at a minute-by-minute scale. This task will create a catalogue of databases of epidemiological data across Europe, with extensive meta-data describing the main characteristics of the available information sources. This catalogue will be integrated with a collaborative platform that will be set up for online discussion and exchange of meta-data among the participants.

### **Task 3.2 – Meta-Model Design.**

**Participants:** FFCUL, IGC, ISI.

**Description:** While some of the previously mentioned datasets are freely available on the Web (e.g. WHO Global Health Atlas, Eurostat), they are often scattered in different repositories, cover partial regions of the world and come in different formats, according to different standards and classifications. The project envisions a unified and integrated approach for the management of these resources, with the design and implementation of an Epidemic Marketplace Platform, publicly available on the web. The platform supports the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. It is an on-line social networking site that will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on a Web2.0 approach, users will become active participants, sharing information and data, and collaborating online, rather than being satisfied with a passive information consumer/viewer

role. We envision proposing a simple reference format, which will facilitate the navigation and use of the datasets. Each dataset will come with a metadata file, signalling general metadata for resource management, containing data such as: the title, the date of submission, version, the source of the data and coverage. Moreover, the metadata will include information for a more thorough description of the data included in the dataset, providing a framework for a more specific description, for example, of epidemiologic and geographic data. The Marketplace will support flexible and intuitive tools for navigation and selection of resources. Standard classifications as well as tagging systems proposed by users will be supported.

### **Task 3.3 – Epidemic Marketplace Platform.**

**Participants:** FFCUL, ISI.

**Description:** This task will implement a platform based on the integration of grid technology and publicly available services and software on the web to support the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. The Epidemic Marketplace Platform will be an on-line social networking site that will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on some of the Web2.0 characteristics, users will become active participants, generating information and providing data for sharing, and collaborating online, rather than being satisfied with a passive information consumer/viewer role. More specifically, researchers can use and contribute to the Marketplace in several different ways. They can: (1) use it as a catalogue of data sources containing the metadata describing existing databases; (2) view, download, tag, and comment on the available resources; (3) provide compliant datasets and relevant information; (4) use it as a forum where to publish information about their own data, seek modellers to collaborate with, share and distribute their new findings.

### **Task 3.4 – Evaluation and monitoring of the use of the catalogue and collaboration services.**

**Participant:** FFCUL.

**Description:** This task involves the monitoring of epidemiological data exchanges performed through the mediating services platform. The evaluation will assess not only the coverage of the catalogued resources, but the users' satisfaction with the user interface and integrated collaborative tools made available through the epidemiological marketplace platform. More importantly, the analysis of the collected datasets and their annotations and usage will provide a rich environment for deriving an epidemiology ontology, which will

help further on the integration and communication among the community of epidemiologists.

## 2.2 Progress in the Reporting Period

- *A summary of progress towards objectives, and details for each task the team was involved;*
- *Highlight clearly significant results;*
- *Team publications within the scope of the project (please provide a pdf of the publication if possible);*
- *List of outreach activities (conferences, Invited talks, presentations, workshops, tutorials,);*
- *List of press releases or media coverage, any particular dissemination activity;*
- *unanticipated finding, opportunity etc*

### **Main Activities at FFCUL in the first semester of the third year of the project in WP3:**

1. Mário Silva and Dulce Domingos attended the Epiwork Review in Brussels in March, 2011 – Mário Silva presented Deliverable D3.3 to the EC and gave an overview of the progress and challenges in WP3. The demo of the first functional Epidemic Marketplace prototype with all the anticipated components was presented to the reviewers.
2. **[opportunity]** The LASIGE team established contacts with other teams involved in projects for the development of biomedical ontologies and terminologies. Miquel Porta, author of the popular “Dictionary of Epidemiology” is very interested in our development of an epidemiological meta-model to describe epidemiological datasets and a possible collaboration may happen.
3. Completed evaluation of Integration of Drupal with Fedora Commons, using the Islandora module extension to Drupal and using this software as basis for a new version of the Repository/Forum of the Epidemic Marketplace and decided for a reorganization of the software architecture of the frontend, which will now interface with Fedora Commons through the EM webservice.
4. Complete re-write of the EM web-services, which do not require the Muradora software and support extended search services based on Apache Solr.

5. Corrado Gioanini visited LASIGE for one week in May for planning the integration of the GLEaMviz platform and the Epidemic Marketplace. Jointly developed the basic use-cases involved in this integration.
6. **[outreach]** The LASIGE team published and presented two full papers at the International Conference of Biomedical Ontologies, in July, which was a significant step towards the disclosure of our work to this important community.
7. **[outreach]** Mário J. Silva was invited to participate in a FuturICT Ethics meeting at ETH Zurich, where a group is being formed under the FuturICT FET Flagship initiative to address the issues of privacy and ethics in IT systems manipulating societal data.
8. **[outreach]** Francisco Couto was invited to give a seminar at EBI in April 2011 on biomedical ontologies.
9. Continued design of new access control method to be implemented in the EM.
10. The second version of the Epidemic Marketplace software, EM 2.0, platform has been in the operation since the end of 2011. Major features: new user interface and component integration based on the Drupal Content-Management System, improved access control management, redesigned web-services.
11. **[outreach]** The LASIGE team participated with five of its members in the EE2, Epiwork/Epifor Workshop, with three accepted posters.

#### **Publications and Presentations:**

1. **[presentation]** F. Couto, Exploring the semantics of biomedical ontologies, in External Seminar at European Bioinformatics Institute, Hinxton, UK, April 2011.
2. **[presentation]** Mário J. Silva. Privacy and Crowdsensing: Can't we just be friends? FuturICT's Ethics Meeting, Zurich, June 2011.
3. **[publication + presentation]** Catia Pesquita, Francisco Couto, Where GO is going and what it means for ontology extension. Proceedings of ICBO 2011, International Conference on Biomedical Ontology, July 2011.
4. **[publication + presentation]** Bruno Tavares, Hugo Bastos, Daniel Faria, Joao D. Ferreira, Tiago Grego, Catia Pesquita, Francisco Couto, The Biomedical Ontology Applications (BOA) framework. Proceedings of ICBO 2011, International Conference on Biomedical Ontology, July 2011.

5. **[publication + presentation]** João D. Ferreira, Francisco Couto, Generic semantic relatedness measure for biomedical ontologies. Proceedings of ICBO 2011, International Conference on Biomedical Ontology, July 2011.
6. **[presentation]** Presentation of EPIWORK to students of the Master/PhD in Epidemiology at the Faculty of Medicine of the University of Lisbon, in an invited 3 hours seminar in July 2011, by Mário J. Silva.
7. **[publication]** Francisco Couto, Mário J. Silva. Disjunctive Shared Information between Ontology Concepts: application to Gene Ontology. Journal of Biomedical Semantics. Vol 2(5). doi:10.1186/2041-1480-2-5.
8. **[presentation]** Mário J. Silva. Challenges in Societal Data Management. Keynote presentation at IDEAS'11: 15th International Database Engineering & Applications Symposium. Lisbon, September, 2011.
9. **[publication + presentation]** Juliana Duque, Mediação Dados-Informação: Design de Informação para a Epidemic Marketplace Master Thesis, University of Lisbon, School of Fine Arts, November 2011 (in Portuguese).
10. **[publication + presentation]** Carla Patrícia Freitas Sousa, Epidemic Marketplace: Repositório e Web Services, Master Thesis, University of Lisbon, Faculty of Sciences, January 2012.
11. **[publication]** Zamite, J., Silva, F., Couto, F., Silva, M. 2011: MEDCollector: Multisource epidemic data collector. Transactions on Large-scale Data-and Knowledge-centered Systems IV: Special Issue on Database Systems for Biomedical Applications (6990), pp. 40-72. Springer-Verlag New York, Inc. ISBN 978-3-642-23739-3
12. **[presentation]** Corrado Gioannini, João Zamite, Integrating the Gleamviz Simulator Tool with the Epidemic Marketplace Platform. Poster presented at EE2, Epiwork/Epifor 2nd International Workshop: Facing the Challenge of Infectious Diseases. 2012.
13. **[presentation]** João Zamite, Dulce Domingos, Mário J. Silva, Owner-Centred Group-Based Access Control for Epidemic Resources. Poster presented at EE2, Epiwork/Epifor 2nd International Workshop: Facing the Challenge of Infectious Diseases. 2012. [http://xldb.di.fc.ul.pt/xldb/publications/Zamite.etal:EpidemicGroupBasedAccessControl:2012\\_poster.pdf](http://xldb.di.fc.ul.pt/xldb/publications/Zamite.etal:EpidemicGroupBasedAccessControl:2012_poster.pdf)

14. **[presentation]** João D. Ferreira, Francisco M. Couto, Mário J. Silva, Ontologies in the Epidemiological Domain. Poster presented at EE2, Epiwork/Epifor 2nd International Workshop: Facing the Challenge of Infectious Diseases. 2012.  
[http://xldb.di.fc.ul.pt/xldb/publications/Ferreira.etal:OntologiesInThe:2012\\_poster.pdf](http://xldb.di.fc.ul.pt/xldb/publications/Ferreira.etal:OntologiesInThe:2012_poster.pdf)
15. **[publication]** João D. Ferreira, Catia Pesquita, Francisco Couto, Mário J. Silva. Epiwork Deliverable 3.5: Epidemic Data Ontology. Technical Report. University of Lisbon, Faculty of Sciences, LASIGE, January 2012.
16. Mário J. Silva, Francisco M. Couto, Dulce Domingos, João Ferreira, Paulo Graça, Tiago Posse, Carlos Sousa, João Zamite. Epiwork Deliverable 3.4 Epidemic Marketplace Platform Report at Month 36, Technical Report. University of Lisbon, Faculty of Sciences, LASIGE, January 2012.
17. **[presentation]** F. Couto, Untangling Biomedical Ontologies, Practical workshop: Bioinformatics and Systems Modelling, Faculty of Sciences, University of Lisbon, 2011

#### **Activities at FFCUL in the third year of the project in Task 3.1:**

- The data collection activity started at the end of the first year and continued as before, with datasets being periodically assembled and uploaded into the Epidemic Marketplace.
- With the release of version 2.0 of the Epidemic Marketplace platform, all the datasets and metadata, which had been loaded in the initial prototype have been converted to the updated meta-data format and migrated to the new platform.

#### **Activities at FFCUL in the third year of the project in Task 3.2:**

- In the period, we worked on the development of a proposal of a network of ontologies for representing the data most commonly used in epidemiology, epidemics modelling in particular, reported in the **Deliverable D3.5**, which was completed by the end of the third year.

#### **Activities at FFCUL in the third year of the project in Task 3.3:**

- The second prototype of the Epidemic Marketplace was presented in the March 2011 Project Review in Brussels.



- WP3/WP4 collaboration on the integration of the computational platform with the data platform.
- Public release of the Epidemic Marketplace at <http://epimarketplace.net>. Anyone from the community can now define collections and retrieve/upload datasets.
- Continued design of new access control system to be implemented in the EM.
- EM 2.0 became the production version at the official site <http://epimarketplace.net>. Major features: new user interface and component integration based on the Drupal Content-Management System, improved access control management, redesigned web-services.
- Integration of GleanViz with the Epidemic Marketplace. We jointly developed an interface that will enable GleanViz upload simulation results as datasets to the EM platform to be shared by the community.

#### Activities at FFCUL in the third year of the project in Task 3.4:

- Continued work on planning of the monitoring and log data collection and analysis tasks. Progress on this domain is documented in Deliverable D3.4.
- Initiated planning on how to use the resources, especially their meta-data, for inferring relationships among the ontological concepts (this is addressed in Deliverable D3.5).

## 2.3 Effort Allocation

The effort allocated by LASIGE to WP3 in the reporting period (and previous years) is as follows:

<b>WP3</b>	<b>FFCUL</b>	<b>Effort</b> (p.m)
<b>Reporting Period</b>		Non-Perm + Perm =Total
Year 1: February 1, 2009 to January 31, 2010		19.38 + 4.53 = 23.91
Year 2: February 1, 2010 to January 31, 2011		32.55 + 6.43 = 38.98
<b>Year 3: February 1, 2011 to January 31, 2012</b>		<b>40.86 + 8.39 = 49.26</b>

The commitment to the project in the first year was 23.91 persons.month, 19.38 p.m from technicians, and 4.53 p.m from permanent staff.

The commitment to the project in the second year was 38.98 persons.month, 32.55 p.m from technicians, and 6.43 p.m from permanent staff.

***In the reporting period (3<sup>rd</sup> year), the commitment was 49.26 persons.month, 40.86 pm from technicians, and 8.39 pm from permanent staff.***

In the third year, it is visible a significant increase in the manpower involved in software implementation tasks, which is in line with our plan and reflects that we are now having a thorough understanding of the required functionalities of the Epidemic Marketplace and how it should interface with Epiwork's Computational Platform.

At present, the FFCUL/LASIGE commitment is now, at the 3<sup>rd</sup> year, above 150% of the manpower planned for the entire duration of the project in this work package. The effort allocated/year is now about 4 times above planned (assuming a 10.25 p.m constant effort/semester), and reflects the decision of hiring more (and less skilled) staff for the FFCUL/LASIGE team of Epiwork. On the other hand, it also reflects an additional commitment by FFCUL/LASIGE to the project to address the weaknesses resulting from having hired less skilled researchers and the need to add more developers to the team.

## 2. Work in WP4 — Epidemic Modelling Platform

This task is lead by ISI, with a total FFCUL contribution of 19 persons.month (12 hired + (7) academics).

### **Activities at FFCUL in third year of the project in WP4:**

The LASIGE work on WP4 is related to the integration of the Computational and Data Platforms of Epiwork, which has been accounted in detail in the description of activities of LASIGE on WP3. Below, we mention the most significant of these activities:

- Mário Silva and Dulce Domingos participated in the March 2011 project review in Brussels.
- Corrado Gioanini visited LASIGE for one week in May for planning the integration of the GLEaMviz platform and the Epidemic Marketplace. João Zamite and Corrado Gioanini developed the basic use-cases involved in this integration. Discussion of the requirements of each use-case and the planned implementation including Communication's Requirements - protocols, authentication and web-services - and Access Control Requirements - based on the Group-Based approach currently being developed.
- João Zamite visited ISI for one week in October for detailing the implementation plans for the integration of the GLEaMviz platform and the Epidemic Marketplace. A mechanism for Distributed Authentication and Access control based on LDAP and the access control framework was designed and is now supported both in the EM and GLEaMviz.
- By the end of the reporting period, the ISI and FFCUL team were actively working on the optimization of the software that implements the API supporting the protocols for interconnecting GLEaMviz with the Epidemic Marketplace.

### **2.4 Effort Allocation**

The effort allocated to WP4 in the in the reporting period (and previous years) is as follows:

<b>WP4</b>	<b>FFCUL</b>	<b>Effort (p.m)</b>
<b>Reporting Period</b>		Non-perm + Perm =Total
Year 1: February 1, 2009 to January 31, 2010		5.72+0.50 = 6.22
Year 2: February 1, 2010 to January 31, 2011		5.05 + 0.40 = 5.45
<b>Year 3: February 1, 2011 to January 31, 2012</b>		<b>7.31 + 2.53 = 9.84</b>

The commitment to the project in the first year was 6.22 persons.month, 5.72 p.m from technicians, and 0.50 p.m from permanent staff.

The commitment to the project in the second year (the reporting period) was 5.45 persons.month, 5.05 pm from technicians, and 0.40 p.m from permanent staff.

***In the reporting period (the 3<sup>rd</sup> year), the commitment was 9.84 persons.month, 7.31 pm from technicians, and 2.53 pm from permanent staff.***

This is as planned, reflecting the fact that this task is led by ISI and our contribution is intensifying in the second half of the project, as we settle on a common software architecture, including interfaces, for orchestrating and running services on both platforms.

The reported effort by technicians reflects the splitting of the effort dedicated to setting-up the hardware and base software of the Epiwork infrastructure in Lisbon between WP3 and WP4 in the first year, and, in the second and third years, the development of the SimpleEMClient for synchronizing local folders with EM streams and work on demonstrating the integration of GleanViz with the Epidemic Marketplace in cooperation with ISI.

## Work in WP7 — Management

This task is lead by ISI.

FFCUL Effort in this task: 4 persons.month.

### Activities at FFCUL in the fifth semester of the project in WP7:

1. Data collection activities for project tracking, preparation of the 5<sup>th</sup> semester report and 3<sup>rd</sup> year reports, management presentations.
2. Mário Silva and Dulce Domingos attended the March 2011 Project Review in Brussels and the third Epiwork Meeting in Courmayeur, Italy in January 2012.

## 2.5 Effort Allocation

The effort allocated to WP7 in the in the reporting period (and previous years) is as follows:

<b>WP7 Reporting Period</b>	<b>FFCUL Effort (p.m) Non-Perm + Perm = Total</b>
Year 1: February 1, 2009 to January 31, 2010	0.0 + 1.10 = 1.10
Year 2: February 1, 2010 to January 31, 2011	0.0 + 0.96 = 0.96
<b>Year 3: February 1, 2011 to January 31, 2012</b>	<b>0.0 + 0.99 = 0.99</b>

The effort dedicated to the project in the period, 0.99 p.m., was 100% contributed by permanent staff, as planned.